

Temporal Reliability of Contingent Behavior Trip Data in Kuhn-Tucker Recreation Demand Models

Lusi Xie

Postdoctoral Researcher, University of Delaware. 265 Townsend Hall, Newark, DE, 19716, USA.
lxie@udel.edu

Wiktor Adamowicz

Professor, University of Alberta. 501 General Services Building, Edmonton, Alberta, Canada.
T6G2H1. Vic.Adamowicz@ualberta.ca

Abstract

Contingent behavior (CB) trip data, eliciting intended trip decisions with hypothetical scenarios, has been popular in recreation demand models. Unlike other stated preference methods, the temporal reliability of CB data has not been examined in recreation demand models, especially in a Kuhn-Tucker (KT) framework. This paper assesses the temporal reliability of CB trip data collected over three years in KT models. We find that coefficient and welfare estimates are largely reliable over time. Our findings add confidence in using CB trip data to model demands within and beyond recreation contexts and provide insight into the broader application of KT models.

Keywords: Contingent Behavior, Stated Preference, Reliability, Recreation Demand

JEL Codes: C83, Q26, Q51

Appendix materials can be accessed online at:

<https://uwpress.wisc.edu/journals/pdfs/LE-99-2-Xie-appA.pdf>

<https://uwpress.wisc.edu/journals/pdfs/LE-99-2-Xie-appB.pdf>

<https://uwpress.wisc.edu/journals/pdfs/LE-99-2-Xie-appC.pdf>

<https://uwpress.wisc.edu/journals/pdfs/LE-99-2-Xie-appD.pdf>

<https://uwpress.wisc.edu/journals/pdfs/LE-99-2-Xie-appE.pdf>

<https://uwpress.wisc.edu/journals/pdfs/LE-99-2-Xie-appF.pdf>

1. Introduction

Travel cost recreation demand models analyze individuals' decisions in outdoor recreation activities using micro-econometric frameworks and revealed preference (RP) data. As one of the non-market valuation methods, recreation demand models provide measures of economic values (i.e., use values) for the environmental amenities of recreation sites and inform policy evaluation, resource management, and damage assessment (Phaneuf and Smith 2005). To address some challenges with RP data, researchers have sought out new data sources, in particular forms of stated preference (SP) surveys (Adamowicz, Louviere, and Williams 1994; Bertram et al. 2020). Originally proposed by Englin and Cameron (1996), contingent behavior (CB) trip data, have become a popular type of SP data that complement RP data in recreation demand models (Bertram et al. 2020; Nobel et al. 2020; Yi and Herriges 2017; Abbott et al. 2018). As a type of SP data, CB data are usually collected by surveys where respondents indicate intended behavior in quantities or frequencies (e.g., units of products to purchase, number of recreational trips to take) under hypothetical scenarios. Different from other SP methods such as contingent valuation (CV) and choice experiment (CE) methods that focus on value elicitation, CB methods focus on eliciting behavior. Nevertheless, due to their hypothetical nature, the accuracy of CB data and estimates from models that use CB data requires investigation.

Given that “true values” are unobserved, the accuracy of the welfare estimates from a non-market valuation method is usually measured by validity (unbiasedness) and reliability (minimum standard errors) (Bishop and Boyle 2017). Validity is often assessed in three aspects (Bishop and Boyle 2019): criterion validity that compares SP estimates with a criterion that involves real money payments, content validity that examines the research procedures qualitatively, and construct validity that compares the convergence of SP estimates with theoretical predictions or other

empirical results. Convergent validity, a special case of construct validity, compares SP estimates with estimates from other methods (especially RP data). Reliability assessments include all procedures that in a study that affect the magnitude of a variance in addition to choosing the estimator that minimizes variance (Bishop and Boyle 2019). In practice, reliability is mostly investigated in test-retest experiments where estimates obtained from two or more time points using the same research procedure are compared. In recreation demand studies, the reliability of coefficient and welfare estimates is usually examined with RP data or combined RP-CB data over several time periods and is considered as assessing temporal convergent validity in some studies. Ji, Keiser, and Kling (2020) test the reliability of welfare estimates in repeated discrete choice models (DCM) with a five-year panel of RP trip data. They find welfare estimates associated with changes in water quality are not reliable across all time periods while welfare estimates of site closures are more reliable. Using the same data, Yi and Herriges (2017) examine and find reliability (i.e., convergent validity) of estimates in DCM with RP data from two consecutive years. Meanwhile, with combined RP and CB trip data, Yi and Herriges (2017) and others (e.g., Jeon and Herriges 2010; Whitehead et al. 2010; Grijalva et al. 2002) examine temporal convergent validity of estimates following the same reliability concept. Although findings on temporal reliability or convergent validity in these studies vary by data collection approaches (e.g., time intervals, RP and/or CB data) and model specifications (e.g., whether to include Alternative Specific Constants, ASC), temporal reliability of estimates from recreation demand models with only CB trip data has not been examined to the best of our knowledge. In addition, model specifications used in the existing studies mostly focus on discrete choice models.

The objective of this study is to assess the temporal reliability of estimates in Kuhn-Tucker (KT) recreation demand models with contingent behavior (CB) data. By adding variation through

randomly assigning changes to recreation site attributes, CB data potentially address identification issues that arise in model estimation with only RP data (von Haefen and Phaneuf 2008; Yi and Herriges 2017), and therefore have been mostly combined with RP data. Unreliable CB data would cast doubt on estimates from models that combine RP and CB data or studies on convergent validity of RP and CB data, and raise concerns about CB question design. At the same time, by collecting intended behavior with hypothetical policy scenarios, CB data can help construct ex-ante welfare estimates for policy analysis. Reliable welfare estimates with CB data can shed light on policies that may take several years to be in place. Compared to the commonly used repeated DCM (Lupi, Phaneuf, and von Haefen 2020), the KT recreation demand model with a multiple discrete-continuous extreme value (MDCEV) specification (Bhat 2008) has several advantages such as assumptions on choice occasions, error term structure and parameters to capture substitution behavior (Bhat 2008; Lloyd-Smith et al. 2020).¹ However, its application has been limited due to computational challenges in the early versions of software, difficulty of parameter interpretation, as well as requirements on sample size.² Applications of KT models with MDCEV specifications have focused on predicting behaviors, recasting choice sets, and capturing seasonal and substitution behavior (e.g., Abbott and Fenichel 2013; Lloyd-Smith et al. 2020). Yet no studies that we are aware of have evaluated the temporal reliability of estimates from KT models in a MDCEV specification with datasets from multiple years.³ As such, this paper focuses on the temporal reliability of estimates from KT models with CB data.

The contingent behavior data in this study are collected from three surveys in 2018, 2019, and 2020. Three distinct samples of recreational hunters in Alberta, Canada indicate intended trip decisions (where to take hunting trips and how many trips) with hypothetical scenarios that propose policy programs to encourage hunting for controlling a wildlife disease, Chronic Wasting

Disease (CWD). Utilizing the discrete and continuous characteristics of the trip data, we estimate separate KT models in the same MDCEV specification with the three CB datasets. We construct welfare estimates of site closures and cross-price elasticities with parameter estimates and underlying data in a simulation-based approach. We test the temporal reliability of parameter, welfare and elasticity estimates over the survey periods of 2018 to 2020 (annual comparisons of 2018 and 2019, 2018 and 2020, and 2019 and 2020).

The results show that more than half of the coefficient estimates, including coefficients capturing individuals' preferences towards the wildlife disease and policy programs, are temporally reliable across the survey years. Welfare estimates of site closures for individuals who take trips to the corresponding sites are largely temporally reliable. Cross-price elasticities of demand for the numeraire good with respect to increased travel costs are similar but not temporally reliable. The mostly reliable estimates indicate that individuals would be more likely to take hunting trips to help control the wildlife disease when responding to policy programs in targeted areas with high disease prevalence. These areas also have a larger use value on average for individuals. Robustness checks with the same KT models using combined RP and CB data also support these findings.

This study contributes to the literature of non-market valuation and recreation demand in the following aspects. Following the standard reliability test procedures (Bishop and Boyle 2017), we extend the test framework in the previous studies (e.g., Ji, Keiser, and Kling, 2020) by considering several aspects of temporal reliability, from coefficient estimates to welfare estimates of site closures to cross-price elasticities of demand with increased travel costs. We also use likelihood ratio tests and non-parametric tests for testing temporal reliability. Our approach shows that this framework can be applied to not only RP data, but also CB data and RP-CB data, and therefore have the potential to be used and extended for other recreation demand models or other topics that

use RP, CB, RP-CB data. Our estimates with CB trip data suggest that CB trip data are a reliable data source in non-market valuation to construct use values, and to compare and combine with RP data in recreation demand models. By evaluating the temporal reliability of estimates from KT models with a MDCEV specification, we provide insights into a broader application of this particular KT model. Previous recreation demand studies focus largely on advancing repeated DCM with few studies on KT models (Lupi, Phaneuf, and von Haefen 2020; von Haefen and Phaneuf 2005). The KT model with MDCEV specification has not been widely used and the temporal reliability of its estimates has not been examined due to computational challenges. This study shows that it is promising to apply KT models with multiple years of data at different sample sizes. Our study also provides implications for policy decisions that use estimates from recreation demand models. In general, reliability tests of estimates with CB data share the same significance for benefit transfer over time in policy analysis as the reliability tests of RP estimates discussed in Ji, Keiser, and Kling (2020). In particular, the temporal reliability of estimates in our empirical application provides confidence in policy advice for wildlife managers on managing a wildlife disease that is slowly progressive and requires management efforts over a long time period.

In the following sections of the paper, we first introduce the background of the empirical application in Section 2. Then we describe the datasets from three surveys we use in Section 3. In Section 4, we introduce the Kuhn-Tucker recreation demand model and reliability tests in the analysis. In Section 5, we report parameter and welfare estimates from the Kuhn-Tucker model as well as results of reliability tests. This is followed by conclusions in Section 6.

2. Background: Chronic Wasting Disease (CWD) and Recreational Hunting

Testing temporal reliability of estimates with contingent behavior data in this study not only has its importance in the recreation demand literature but also has direct policy implications for management of a wildlife disease – Chronic Wasting Disease.

Chronic Wasting Disease (CWD) is a fatal prion disease that is infectious among farmed and wild deer populations. As of 2022, CWD has affected 4 Canadian provinces, 28 states in the United States, 3 European countries, and in South Korea.⁴ While CWD has not been found to transmit from animals to humans, health agencies raise the concern of this potential and recommend avoiding using and consuming infected animals.⁵ In Alberta, CWD has affected wild deer, especially mule deer since 2005. While the CWD prevalence remains relatively low, CWD has become more prevalent and spread to a larger geographic area over the recent years (Pattison-Williams et al. 2020) – this raises the concern about the reduction in wildlife populations if the prevalence is high enough as in areas like Wyoming (DeVivo et al. 2017).

Although CWD is challenging to control, wildlife agencies have designed different CWD surveillance and management programs. Currently, the only feasible control approach is reducing animal populations in infected regions. Over the past few years, engaging recreational hunters in depopulation has become desirable and wildlife agencies in western North America have implemented or proposed some incentive programs to increase hunter harvests in CWD-infected areas (Cooney and Holsman 2010; Holsman and Petchenik 2006; Holsman, Petchenik, and Cooney 2010; Western Association of Fish and Wildlife Agencies 2017). As recreational hunters obtain use values by taking hunting trips and harvesting animals, incentive programs could increase their hunting opportunities in CWD-infected areas. On the other hand, recreational hunters who consume meat from harvested animals and perceive CWD as a risky disease might be less satisfied

if some harvested animals are CWD-positive. Therefore, it is important to examine hunters' responses to incentive programs and use values associated with hunting activities before incentive programs are implemented. CB responses provide an ideal data source for modeling hunters' intended trip decisions with proposed incentive programs.

Limited effort has been made to understand hunting behavior over time, although it is important and even required for engaging hunters in CWD management. Since CWD is a slowly progressive disease, sustained efforts are required to reduce the prevalence and slow the spread for CWD control. Accordingly, from an epidemiological perspective, the effectiveness of CWD management programs can only be evaluated when the programs are implemented persistently for a certain time period, e.g., a minimum of 5 years as recommended by Western Association of Fish and Wildlife Agencies (2017). However, most previous CWD management programs were implemented for a short period of two to three years or were not implemented continuously in a longer time period (Conner et al. 2007; Western Association of Fish and Wildlife Agencies 2017) – this limits the understanding of whether these management programs could control the disease over the long time. Moreover, hunters' opinions and behavioral responses to CWD and associated incentive programs might change as CWD evolves over time. Yet researchers and policy makers have not paid much attention to continuously collecting data from hunters to evaluate the impacts of CWD and programs on hunters (Vaske and Lyon 2011; Cooney and Holsman 2010; Holsman and Petchenik 2006), regardless of the status of incentive programs (e.g., whether the programs have been discontinued, are being implemented, or are under discussion). As such, it is not clear whether and how incentive programs could encourage hunter harvests to control CWD over time.

In Alberta, CWD prevalence and spread are slowly increasing (i.e., CWD is not newly discovered and is familiar to hunters); hunter population sizes are relatively stable; and wildlife managers are

increasingly interested in designing incentive programs for CWD control. As a result, we have the opportunity to assess the temporal reliability of estimates from modeling hunting behavior to provide policy advice.

3. Data

To test the temporal reliability of estimates, we use a pseudo panel dataset from three surveys administered to distinct samples from 2018 to 2020. In this section, we describe these surveys and contingent behavior questions, as well as summary statistics of the data.

Surveys and contingent behavior questions

We administered three surveys to distinct samples of recreational hunters in Alberta on Qualtrics in 2018, 2019, and 2020 respectively. The surveys were sent out roughly at the same time period (between February and May) that is after the hunting season in November of the previous year (2017, 2018, 2019). The target population for the surveys was recreational hunters who held special licenses for mule deer in CWD-infected and surrounding areas as with the dark grey boundaries in Appendix Figure A1. For each survey, 5,000 eligible individuals were randomly drawn from the license database of Alberta Environment and Parks. In 2018 and 2020, one invitation and one reminder were sent out to eligible participants by Alberta Environment and Parks on our behalf. In 2019, only one invitation email was sent out because the reminder email was cancelled due to the provincial election in Alberta. After excluding respondents who did not agree to participate in the survey, did not take hunting trips in the previous year, or did not provide required information (e.g., hunting trips, postal codes for travel cost calculation), we construct a pseudo panel dataset that consists of data from 636, 330, and 873 respondents in each year. As we

did not collect unique identifiers from respondents (i.e., Wildlife Identification Number), we are not able to identify if we have same individuals across years.

Following a section that collects RP data, the three surveys include a section that consists of four contingent behavior scenarios to collect SP data as summarized in Table 1. These CB scenarios are identical across years. Each CB scenario proposes an incentive program that aims to increase hunting trips for CWD management. Incentive programs proposed in the scenarios are based on policy recommendations in Western Association of Fish and Wildlife Agencies (2017) and discussions with wildlife managers and biologists from the Government of Alberta. Two scenarios propose to extend hunting seasons in October or December from the regular season in November in sampling areas of 65 hunting sites. Season expansion programs are recommended to reduce prevalence and slow the spread of CWD and thus apply to CWD-infected and surrounding areas (Western Association of Fish and Wildlife Agencies 2017). The other two scenarios provide material incentives of either extra tags or monetary rewards in gift cards in the regular season of November in 11 high CWD prevalence hunting sites. Extra tags or gift cards are to provide incentives to increase harvest at targeted areas where CWD prevalence is higher than 10% as of 2016 (Western Association of Fish and Wildlife Agencies 2017). Season expansion and extra tags/gift cards, either being implemented independently or together, are considered to be effective to curb CWD (Western Association of Fish and Wildlife Agencies 2017). As such, we design four scenarios that cover all aspects of recommended policy programs.

[[Insert Table 1 here]]

Each respondent randomly received two CB scenarios in the three surveys. Some respondents in the 2018 survey received two season expansion scenarios whereas all respondents in the 2019 and 2020 surveys received one scenario with season expansion (either October or December) and one

scenario without season expansion (either extra hunting tags or gift cards). After presenting CB scenarios to respondents, the surveys asked how many trips they would have taken in the previous hunting season (and extended seasons with October and December scenarios) – the same time as their actual hunting trips (see Appendix F for screenshots of CB scenarios in the surveys). We asked the questions in a retrospective manner (i.e., what they would have done) rather than in a forward-looking manner (i.e., what they will do) so that respondents would be more likely to hold all other factors constant when responding.

As discussed in Ji, Keiser, and Kling (2020), one cause of unreliable coefficient and welfare estimates over time is changing preferences (i.e., unstable preferences) that could be captured by utility parameters. Changes in preferences might arise from external shocks, such as policy changes, financial crises and pandemics. In Alberta, there was a policy change on hunting licenses associated with CWD between 2018 and 2019 surveys. Prior to 2018, recreational hunters could obtain a free replacement license if they chose not to consume meat from CWD-infected animals. This program was discontinued from the 2018 hunting season and therefore potentially affected trip decisions collected in 2019 and 2020 surveys.⁶ Although the 2020 survey was implemented in March 2020, during the COVID-19 pandemic, the primary data collection after the initial invitation happened before the first 30 cases were found in Alberta. In addition, there were no stringent stay-at-home or lockdown orders in place when the 2020 survey was in the field. As such, we are not aware of external shocks that could affect preferences from 2018 to 2020 surveys except for the discontinued replacement license program.

Descriptive statistics and trip data

Table 2 presents descriptions and mean values of main variables in three surveys. The only site attribute is CWD prevalence rate that is calculated as the percentage of positive CWD cases in

mule deer over the total number of mule deer heads submitted for testing from hunters. As CWD testing results usually come after the hunting season, recreational hunters only have CWD information from the previous hunting season when they make hunting trip decisions. Therefore, we use the CWD prevalence rate from the previous hunting seasons in 2016, 2017, and 2018. As we can see, CWD prevalence has increased on average over the survey period. Moreover, CWD has spread to a larger area, i.e., from west to the east in Alberta, as shown in Appendix Figure A1. Since CWD prevalence is the only available attribute that varies by site and year, we use the actual prevalence rate rather than converting it to a categorical measure as in Zimmer, Boxall, and Adamowicz (2012) and Xie, Adamowicz, and Lloyd-Smith (2020) where they only use cross-sectional data in one year.

[[Insert Table 2 here]]

CB scenario dummy variables identify the impacts of policy programs that vary by sites and individuals. As shown in the third column of Table 1 and the early discussion, the four scenarios do not apply to the same areas: October and December season expansion programs apply to sampling areas while extra tags and gift cards programs only apply to areas with high CWD prevalence. Although respondents were not restricted to indicate intended trips only to eligible areas, the CB scenario dummy variables are defined to distinguish the different impacts of policy programs by sites. At the same time, respondents randomly received two scenarios out of four, CB scenario dummy variables also identify who received what scenarios and therefore are different across individuals. Given the definitions of these dummy variables, one should be cautious about interpreting these variables because the reference category when dummy variables are equal to 0 is not a specific policy program, but an indication when and where associated policy program does not apply – either a site is not in the eligible area of the policy program or an individual did not

receive the scenario.⁷ The extended season dummy variable is to distinguish trip decisions during the regular season in November with trip decisions during the extended season in October or December.

Socio-demographic variables are not balanced across surveys. Although the sampling methods were the same for the three surveys, the participation was voluntary and there might have been self-selection bias issues in responses that are similar over the three years. Variables such as age and income have similar distributions across years and there is no systematic difference from one year to another. Given that our sampling was exogenous from recreational trips, and we do not have the necessary information on the population of our interest (i.e., active mule deer hunters in the study area), we do not construct sampling weights and conduct weighted estimation (Lupi, Phaneuf, and von Haefen 2020). We also calculate the round-trip travel costs that consist of out-of-pocket monetary expenses and opportunity costs of travel time. We convert the travel costs in 2017 Canadian dollars as the first hunting season we have information on was in 2017.

Appendix Table A1 presents the average number of trips that were reported by each person (RP) and would have been taken by a person with CB scenarios in each survey. As some responses are not eligible based on the criteria listed in the previous section, the number of respondents in each scenario is not the same. The average number of trips under the most scenarios in the 2018 survey is slightly lower than those in 2019 and 2020 surveys – this pattern is consistent with their parallel RP trips in the 2017, 2018, and 2019 hunting seasons. Since October and December season expansion programs provide longer hunting seasons, the average numbers of trips in these two scenarios are higher than the average numbers of trips in extra tags and gift cards. The same reason explains a higher average number of trips in December compared to October. We further break down the average number of trips per person per site by targeted and non-targeted areas for each

scenario in each year in Appendix Table A2. Compared to RP trips, we find that more CB trips would have been taken with October and December extended seasons to both targeted and non-targeted areas because season expansion relaxes the time constraints on the recreation trips and allows for temporal substitution from the regular season to the extended season. Whereas with extra tags and gift cards scenarios, more trips would have been taken to the targeted area instead of non-targeted area because extra tags and gift cards are only provided to trips in the targeted area without relaxing the time constraints.

4. Analysis

In this section, we outline analysis methods we use to test for reliability. Following the standard practice of reliability tests (Bishop and Boyle 2017) and extending the framework by Ji, Keiser, and Kling (2020), we first estimate a recreation demand model separately for each survey. Then we construct welfare estimates of site closures and cross-price elasticities. To assess reliability of coefficient and welfare estimates, we will test for 1) differences in the coefficient estimates from the same model across three years; 2) differences in the associated welfare estimates and cross-price elasticities across three years. Given that CB data are usually collected following RP data, the same sets of analyses are conducted with combined RP and CB data as robustness checks.

Model estimation

As shown in Appendix Table A1, respondents would have taken more than 5 trips in all scenarios across years. In order to make use of the “continuous” nature of count data while accounting for potential zero trips, we apply a KT model with the multiple discrete-continuous extreme value (MDCEV) specification (Bhat 2008), for its advantages over a repeated discrete choice model and a traditional Kuhn-Tucker model with an LES specification (Lloyd-Smith et al. 2020; Bhat 2008).

The conceptual framework for the Kuhn-Tucker model starts from a constrained utility maximization problem. Each recreational hunter is assumed to maximize utility $U(x_j, Q_j, z)$ – a function of recreation hunting trips x_j to hunting site j , vectors of site attributes Q_j at site j , and a numeraire good z – by choosing the number of recreation trips and consumption of the numeraire good, subject to their budget and time constraints:

$$\max_{x_j, z} U(x_j, Q_j, z) \quad [1]$$

$$\text{subject to } \sum_j p_j x_j + z \leq \bar{y} + t_w w \quad [2]$$

$$\sum_j t_j x_j + t_w \leq \bar{T} \quad [3]$$

Equation [2] is the budget constraint where p_j is the monetary cost of a hunting trip, \bar{y} is the non-wage income, t_w is the time spent on working at parametric wage and w is the wage rate. Equation [3] is the time constraint where t_j is the travel time of a hunting trip that does not include on-site time and \bar{T} is the total available time to the hunter. Following the common practice in travel cost recreation demand models (Bockstael and McConnell 2007), we collapse the two constraints into one as follows:

$$\sum_j (p_j + t_j w) x_j + z = \bar{y} + w \bar{T} \quad [4]$$

The associated first order Kuhn-Tucker conditions of the maximization problem is

$$\frac{\partial U / \partial x_j}{\partial U / \partial z} \leq p_j + t_j w, j = 1, \dots, J \quad [5]$$

$$x_j \left[\frac{\partial U / \partial x_j}{\partial U / \partial z} - p_j - t_j w \right] = 0, j = 1, \dots, J \quad [6]$$

Based on these two conditions, we specify a utility function $U(x_j, Q_j, z)$ and calculate the travel cost $p_j + t_j w$ to derive estimating equations for empirical estimation. The round-trip travel cost,

consisting of the monetary expenses and opportunity cost of travel time (measured in hours), is given by

$$\text{Travel cost} = 2 \times \text{travel distance} \times \text{cost per kilometer} + 2 \times \text{travel time} \times \text{wage rate} \times \frac{1}{3} \quad [7]$$

For the utility function, we choose the translated generalized constant elasticity of substitution (tCES) specification from Bhat (2008):⁸

$$U(x_j, Q_j, z) = \sum_j \frac{\gamma_j}{\alpha} \varphi_j \left[\left(\frac{x_j}{\gamma_j} + 1 \right)^\alpha - 1 \right] + \frac{1}{\alpha} z^\alpha \quad [8]$$

where γ_j , α are utility parameters to allow for the corner solution, satiation effects, and diminishing marginal utility of additional trips or numeraire good as detailed in Bhat (2008). φ_j is the baseline marginal utility of a recreation trip to site j when no trips are taken. Similarly, φ_z captures the baseline marginal utility of the numeraire good when it is not consumed. We further specify $\varphi(Q_j, \varepsilon_j) = \exp(\beta'Q_j + \varepsilon_j)$ for hunting trips and $\varphi_z = \exp(\varepsilon_z)$ for the numeraire good. Q_j includes CWD prevalence, CB scenario dummy variables, and socio-demographic variables in Table 2 (except for income and travel costs). As CWD is the only available site attribute, we also include year invariant alternative specific constants (ASC) for each site to capture the specific preferences towards certain sites and address potential omitted variable bias (Murdock 2006). The error terms ε_j and ε_z are to capture unobserved heterogeneity across individuals. The error terms are assumed to follow a type 1 extreme value distribution with a scale parameter σ and are independent across individuals and choice occasions.

The same specification is used to estimate models using data from 2018, 2019, and 2020 surveys. 79 hunting sites would have been visited in CB responses of 2018 and 2020 surveys while only 72 hunting sites would have been visited in CB responses of the 2019 survey. Therefore the choice

set is slightly different across years, i.e., $J = 79$ for 2018 and 2020 surveys, $J = 72$ for 2019 survey. Note although October and December season expansion scenarios have two time periods (i.e., regular and extended seasons) for individuals to take hunting trips and they might substitute across time, we assume they treat the two time periods independently and use the extended season dummy variable to distinguish it from the regular season. We do not combine hunting site and time periods in the choice set as in Xie, Adamowicz, and Lloyd-Smith (2020) that only use RP and CB data associated with season expansion programs from the 2018 survey, to better compare extra tags and gift card scenarios that do not have the extended seasons.⁹ The model is estimated with the R package `rmdcev` (Lloyd-Smith 2021), using Maximum Likelihood and 50 multivariate normal draws to compute standard errors.¹⁰

Welfare and elasticity simulation

Welfare

Following Lloyd-Smith (2018), we construct welfare measures of site closures by using a simulation approach with model estimates and underlying data used in estimation. The approach first simulates Hicksian demand for each site and then uses the Hicksian demand to calculate the Hicksian compensating surplus CS^H . The CS^H for a price change from baseline levels p^0 to new levels p^1 using an expenditure function is given by (Lloyd-Smith 2018):

$$CS^H = y - e(p^1, U^0, \theta, \varepsilon) \quad [9]$$

where y is the annual income, θ is the vector of utility parameters $(\varphi_j, \alpha, \gamma_j)$, U^0 is the baseline utility level specified as $U^0 = V(p^0, y, \theta, \varepsilon)$, and ε is the error term that captures unobserved heterogeneity by individuals (see Lloyd-Smith 2018 for a full description). In our recreation demand models, p^0 and p^1 are the baseline and new travel costs. In order to simulate CS^H of site

closures, p^1 is set to a very large number so that the new price is much higher than the choke price and therefore essentially has the same effect as site closures (Lloyd-Smith 2022). We consider closing one site at a time for the three surveys and that gives us 79, 72, and 79 policy scenarios for 2018, 2019, and 2020 surveys respectively. We draw 50 conditional errors per individual in each sample to simulate $E(CS^H)$ in each policy scenario.

The direct output of welfare simulation is $E(CS^H)$ of per individual in each policy scenario. As this does not account for differences in trips and visitation patterns to sites across years (e.g., some sites might be more popular than the other sites in one year but not across years), we further calculate welfare estimates per trip, averaging across all individuals (called “welfare estimates per person”) and across only individuals with positive number of trips to sites being closed (called “welfare estimates per participant” as in Lloyd-Smith 2022) for each policy scenario for each sample in the following steps:

1. For each individual in each simulation, divide the welfare estimates by the positive number of trips or keep the welfare estimates (close to 0) if no trips are taken. This gives us welfare estimates per trip in each simulation.
2. Keep welfare estimates of all individuals for welfare estimates per person while only keeping welfare estimates of individuals with at least one trip for welfare estimates per participant.
3. Calculate the average welfare estimates per trip per person or per participant in each simulation.
4. Obtain the mean, 95% confidence interval (low and high) of the welfare estimates per trip per person or per participant for each policy scenario by taking the average, 2.5% and 97.5% quantiles of the average welfare estimates across simulations.

These steps give us three welfare estimates per trip per person or per participant for each policy scenario in each year: mean, lower and higher bounds of the 95% confidence interval. Welfare estimates per person assume that all sites are valuable to the whole sample, regardless whether they visit the sites or not, whereas welfare estimates per participant focus on respondents who would visit the sites. As such, welfare estimates per participant accounts for different visitation patterns across years.

Elasticity

In addition to welfare estimates which capture the “normative” aspect of the model, we also consider the positive/predictive aspect of the model reliability by looking at cross-price elasticities (e.g., demand elasticities to travel cost changes).¹¹ As the KT model captures substitution across sites, in principle, the cross-price elasticities of demand should be calculated for each site across years. This would result in a very large number of comparisons in our case, making the comparisons across years and reliability tests challenging to interpret. To obtain meaningful comparisons, we increase the travel costs to all sites by 10% and calculate the resulting changes of demand for the numeraire good and obtain the corresponding cross price elasticities for each sample in the following steps:

1. For each individual i , calculate the original numeraire good using the formula below

$$z_i^0 = income - \sum_j travel\ cost_j \times trips_j \quad [10]$$

This is adapted from Equation [4] that collapses the time and budget constraints from the conceptual framework.

2. With the model estimates and underlying data used in the original estimation, we simulate the new demand of the numeraire good z_i^1 by increasing the travel costs to all sites by 10% for each individual in each simulation.
3. Calculate the changes (in percentages) of the demand for the numeraire good for each individual in each simulation.

$$\Delta z_i = \frac{z_i^1 - z_i^0}{z_i^0} \quad [11]$$

4. Calculate the cross-price elasticities for each individual in each simulation.

$$e_i = \frac{\Delta z_i}{10\%} \quad [12]$$

5. Calculate the medians of the cross-price elasticities across individuals in each simulation to reduce the impacts of individual outliers.

As a result, we obtain one cross-price elasticity for each simulation for us to construct confidence intervals across simulations for reliability tests.

Reliability tests

To test the reliability of coefficient and welfare estimates as well as cross-price elasticities across years, we conduct a non-parametric test of estimate differences. With the matrix of parameter draws from estimation and matrix of welfare and cross-price elasticity estimates in each simulation, we can calculate the differences of coefficient and welfare estimates and cross-price elasticities across each year's model for each draw/simulation as below:

$$\Delta Estimate_{S,k}^{y_1, y_2} = Estimate_{S,k}^{y_2} - Estimate_{S,k}^{y_1} \quad [13]$$

Where *Estimate* is coefficient estimate of each parameter in each draw (i.e., $\Delta Estimate = \Delta Coefficient$) or welfare estimate (i.e., $\Delta Estimate = \Delta Welfare$) per trip per person or per

participant or cross-price elasticity (i.e., $\Delta Estimate = \Delta e$) in each simulation. S is the draw number or simulation number. k is the parameter number for coefficient estimates, $k = j$ is the site in choice sets for welfare estimates, $k = 1$ for cross-price elasticities (i.e., change of demand for the numeraire good with respect to a 10% increase of travel costs to all sites). y_1, y_2 is one combination from $\{(y_1 = 2018, y_2 = 2019), (y_1 = 2018, y_2 = 2020), (y_1 = 2019, y_2 = 2020)\}$.

We then construct a 90% confidence interval for these differences by deleting values at the top and lowest 5% quantile. If the resulting 90% confidence interval contains 0, the coefficient and welfare estimates are reliable. This non-parametric test is similar to the approach used for testing the reliability of welfare estimates in Ji, Keiser, and Kling (2020) where they obtain the differences by bootstrapping procedures. One should note that the maintained hypothesis of the reliability tests in this paper is that preferences are stable because estimates that measure unstable preferences are likely unreliable.¹²

5. Results

Model estimation, welfare and elasticity simulation

Table 3 presents the estimates of selected baseline marginal utility parameters (φ), satiation (α), and scale (σ) parameters from Kuhn-Tucker model for each year. As we estimate one φ (ASC) and γ parameter for each site, a total of 170 parameters are estimated for 2018 and 2020 surveys and a total of 156 parameters are estimated for the 2019 survey.¹³ Appendix B reports all estimates of φ (ASC) and γ parameters. We see that the CWD estimates are negative and insignificant across years – this is consistent with recent findings in Xie, Adamowicz, and Lloyd-Smith (2020) and Pattison-Williams et al. (2020).¹⁴ It indicates recreational hunters are not driven away by the

presence of CWD even though CWD has an increased prevalence rate and CWD has spread across these years. The reason could be that hunters' perceived CWD risks have declined over time, given the fact that CWD has existed in Alberta for more than ten years and its prevalence level remains relatively low (Vaske and Miller 2019). All estimates of CB scenario dummy variables are positive and mostly significant across years, indicating individuals are more likely to take trips when they receive CB scenarios in certain areas. Comparing the magnitude of CB scenario dummy estimates, the extra tags scenario has the largest impact, followed by the gift cards, December, and October scenarios. The popularity of the extra tags scenario is not surprising and also appears in the qualitative responses in surveys. And this is also shown in Table A2 where average numbers of trips per person per site to the targeted area under the extra tag scenario are the largest compared to other scenarios for almost all years (except for the 2018 survey). One additional hunting tag, by allowing additional harvests during the same hunting trips, increases harvesting opportunities without substantially increasing the travel costs. Although qualitative responses in surveys suggest season expansion is more favored than gift cards, our model estimates indicate that gift cards are slightly more preferred than the December season expansion. A plausible explanation could be the gift cards scenario is proposed to a smaller area that is popular among respondents whereas season expansion scenarios are proposed to a larger area that include some less popular sites. As the presence of CWD does not appear to change hunters' trip decisions (on average), gift card scenarios make popular areas more desirable even though these areas have high CWD prevalence. The December scenario is more preferred than the October scenario, mostly due to the overlap of other hunting seasons in October as indicated by open-ended responses in surveys. The negative and significant coefficient of extended seasons shows that individuals are less likely to take hunting trips in extended seasons (October or December) compared to the regular season in

November. Even though most of the socio-demographic variables are not significantly different from zero, we include them in estimation because they are not all balanced across years and some are significant in the models. We also estimate models without socio-demographic variables and the coefficient estimates are similar to our preferred specification with socio-demographic variables.

As CB data were collected after the RP data and respondents were reminded of their actual trips when providing CB responses in the surveys, respondents likely anchored CB responses with RP trips. As such, preferences towards CB scenarios revealed by estimates with CB data only might be confounded with unobserved factors embedded in RP trip decisions. To account for this potential issue, we estimate the same KT models with combined RP and CB data (see results in Appendix Table D1) as a robustness check. Results are similar to what the CB data alone show based on the signs and significance levels of the estimates, except that the estimate of December scenario coefficients are slightly larger than that of gift card scenario coefficients for 2018 and 2019 surveys. Supported by the robustness check, from the main results of coefficient estimates with CB data, we find that recreational hunters are not affected by CWD prevalence levels and are likely to take more hunting trips when they receive incentive programs, especially the incentive programs targeting areas with high CWD prevalence.

[[Insert Table 3 here]]

Table 4 reports selected welfare estimates per trip per person and per participant of closing sites of our main interest – the 11 sites in areas with high CWD prevalence and where all CB scenarios apply. These sites are also popular among respondents because the average welfare estimates in this area are larger than the average in other areas. We calculate the welfare estimates of closing sites for each site in the choice set and report them in Appendix C. From the first half of the table,

we can see that the mean welfare estimates per trip per person are between -\$0.11 and -\$7.4 Canadian dollars across years, with some welfare estimates in 2019 slightly larger than estimates in 2018 and 2020. The biggest welfare loss of \$9.83 per trip in Canadian dollars is from closing WMU 200 for the 2019 sample whereas the smallest welfare loss of \$0.04 Canadian dollars is from closing WMU 730 for the 2020 sample. The magnitude of estimates is more than twenty times larger for those who would have taken trips to these hunting sites as in the second half of the table. The mean welfare estimates have a wider range between -\$42 and -\$130 Canadian dollars across years. The biggest welfare loss of \$241 Canadian dollars is from closing WMU 730 for the 2019 participants and the smallest welfare loss of \$10 Canadian dollars is from closing WMU 728 for the 2019 participants. The different patterns of welfare estimates indicate that the economic significance of site closures depends largely on the variation of visitation patterns (or where respondents would have taken trips to). When we focus on the entire sample (e.g., welfare estimates per trip per person), the welfare loss of site closures is relatively small. However, site closures could result in much larger welfare loss to those who take trips to those sites. For example, WMU 730 might not be very valuable for the entire sample but its value is very high for those who enjoy hunting there or who live nearby. Welfare estimates from the models with joint RP and CB data tell the same story (detailed in Appendix Table D2). Taken together, these welfare estimates suggest that site closures would result in welfare loss, in particular to individuals who take trips to the closed sites rather than the entire hunter populations.

[[Insert Table 4 here]]

With a 10% increase of the travel costs to all sites, the simulated average demand elasticities for the numeraire good are 0.0196, 0.0330, and 0.0208 for 2018, 2019, and 2020 respectively. This indicates that the numeraire good is a substitute for trips when the travel costs increase and

respondents' demand for the numeraire good is inelastic with the changes of travel costs. With joint RP and CB data, the average demand elasticities are similar to the ones with CB data.

Reliability tests

Using the approach described in Section 4.3, we construct the 90% confidence intervals for a total number of 482 $\Delta Coefficient$ (170 pairs between 2018 and 2020, and 156 pairs each between 2019 and the other two years) and a total number of 223 $\Delta Welfare_{S,j}^{y_1,y_2}$ per person or per participant (79 pairs between 2018 and 2020, and 72 pairs each between 2019 and the other two years). We focus on Figures 1-3 that present estimates of CB scenarios in KT model, welfare estimates of closing three sites in different areas, and percentage of temporally reliable estimates across years.

[[Insert Figure 1 here]]

[[Insert Figure 2 here]]

Figure 1 visually presents the mean and 95% confidence interval of estimates of CB scenario dummy variables – our variables of interest – in the baseline marginal utility parameters from KT model estimation. We can see that the ranking of estimates based on magnitudes are consistent across years, even though the estimate of October is not significantly different from 0 in 2019. 95% confidence intervals of the same coefficient estimate all overlap across years. Our non-parametric tests show that coefficient estimates of extra tags and October are not significantly different (i.e., reliable) across the three years. Coefficient estimates of December and gift cards are not significantly different mostly across two years. The only two pairs that are significantly different (i.e., not reliable) from each other are: the coefficient estimates of December from 2018 and 2020 and those of gift cards from 2019 and 2020. Estimates of ϕ parameters including CWD, extended

season, and most social demographic variables are not significantly different across models. Although our variables of interest in ϕ parameters are mostly reliable across years, one should note that these are only a small proportion of all parameters estimated in the models.

Following the idea in Swait and Louviere (1993), we conduct joint tests of equality of all coefficient estimates with CB data only and RP-CB data respectively.¹⁵ We estimate models for each dataset and one model with the pooled dataset, and conducted likelihood ratio tests using log-likelihood values from the models. Comparing with the critical values of the chi square statistics, we reject the null hypothesis that all parameters are equal across three years using the 95% confidence interval (see Appendix E for results). Although the test results indicate that the coefficient estimates are not reliable, one should note that likelihood ratio tests assume that data come from the pooled model and are easily affected by “outliers” (Hensher et al. 1998). Furthermore, since welfare estimates are non-linear transformations of coefficient estimates (Ji, Keiser, and Kling, 2020), findings from the coefficient estimates are not necessarily consistent with the welfare estimates given the different statistical properties (Krinsky and Robb, 1986).

As such, we turn our attention to the reliability of welfare estimates – a focus of economists and policy makers. Figure 2 presents welfare estimates of closing three sites respectively: WMU 151 has high CWD prevalence and is within the area where all CB scenarios apply, WMU 230 has CWD presence and is within the area where only season expansion scenarios apply, WMU 501 does not have CWD and is outside of the areas with CB scenarios. We see that welfare estimates are negative for the closures of all sites but closing WMU 151 with high CWD prevalence results in the largest welfare loss and closing WMU 501 without CWD presence has the smallest welfare impact. This pattern is consistent across years for both welfare estimates per person and per participant. According to non-parametric tests, welfare losses per person of closing WMUs 230

and 501 are mostly significantly different (i.e., not reliable) across years whereas welfare loss per person of closing WMU 151 is not significantly (i.e., reliable) across years. Interestingly, welfare loss per participant of closing all three sites are not significantly different across years. This suggests temporal reliability of welfare estimates are affected by whether different visitation patterns are accounted in welfare estimates calculation (i.e., per person vs. per participant) and how sites being closed are affected by CWD and CB scenarios.

Since we test the reliability of a number of coefficient estimates in KT models and welfare estimates of closing one site at a time, we report the percentage of reliability estimates in Figure 3. We define years of rejection in a similar manner as in Ji, Keiser, and Kling (2020) except that we do not fix a base year for comparisons because we only have three years of estimates. As such, each estimate has three pairs of comparisons: 2018 vs. 2019, 2018 vs. 2020, 2019 vs. 2020, and we have a total of 156 coefficient estimates and 72 welfare estimates available. In our definition, no rejection means estimates are not significantly different across all years, indicating complete temporal reliability; 1 year rejection means that estimates are only significantly different between one pair; 2 years rejection means that estimates are significantly different between two pairs; and 3 years rejection means estimates are not temporally reliable across years. Figure 3 shows that more than 50% of coefficient estimates are completely temporally reliable and no coefficient estimates are completely unreliable across three years. Around 47% of welfare estimates per person are not reliable across two years and 13% are completely unreliable. The pattern changes dramatically for welfare estimates per participant that considers the different visitation patterns in each year: around 71% of welfare estimates per participant are completely temporally reliable, including 9 (out of 11) sites where CB scenarios apply and only 1.4% (1 site) of welfare estimates per participant are completely unreliable across years. Estimates from models with RP and CB

data, although slightly less reliable, follow a similar pattern as shown in Appendix Figure D1: most coefficient estimates (about 50%) and welfare estimates per participant (about 61%) are reliable across three years.

We also test the reliability of three cross-price elasticities. Although only one pair is not significantly different, the differences of the other two pairs have very narrow 90% confidence intervals and are very close to zero. Therefore, the elasticities are also similar across years.

[[Insert Figure 3]]

Large proportions of temporally reliable coefficient estimates and welfare estimates per participant reinforce our findings in the previous section that hunters would be more likely to take trips over years when incentives are offered. More importantly, sustained efforts to control CWD to avoid site closures could constantly avoid potential welfare loss to targeted hunters. In general, temporally reliable estimates in this study suggest that researchers could use CB data in KT models to construct reliable welfare measures. Policy makers, especially wildlife managers in Alberta, could rely on findings from one-time data collection to design incentive programs for future implementations. However, one should note that our findings are affected by our sampling techniques and empirical model specification. Even with the same maintained hypothesis of stable preferences, the same sampling techniques and model specification might yield different findings from reliability tests in a different study context, time period, and study site. Moreover, with our proposed reliability tests that depend on uncertainty around point estimates, sample sizes may also affect test results in that reliability tests are more likely to fail with larger sample sizes (smaller standard deviations) than with smaller sample sizes. Although this is not the finding in our study because most of the coefficient and welfare estimates are significantly different from zero, one should be more cautious and conservative when interpreting testing results especially with more

insignificant estimates from smaller samples. When applying the methods used in this study to a different study to test the temporal reliability of CB trip data in a KT model, one can borrow lessons from studies that discuss benefit transfer across study sites (e.g., Bateman et al. 2011).

6. Conclusion

In this paper we assess the temporal reliability of estimates in Kuhn-Tucker recreation demand models with contingent behavior data. By collecting intended trip decisions in three surveys administered to distinct samples of recreational hunters from 2018 to 2020, we examine how individuals would respond to proposed incentive programs that aim to control a wildlife disease, Chronic Wasting Disease. Making use of the site-specific count trip data, we estimate three KT models with the same specification for each year's data respectively and construct associated welfare measures of site closures. We use non-parametric tests to examine the temporal reliability of coefficient and welfare estimates as well as cross-price elasticities. We also conduct the same sets of analysis with joint RP and CB data as robustness checks. We find that individuals are not driven away by the wildlife disease and they respond consistently to incentive programs over time. Extra hunting tags in targeted areas with high disease prevalence are mostly favored by hunters across years, seconded by gift cards that apply to the same areas. Season expansion programs, by applying to a larger area, have smaller and consistent impacts on trip decisions. Welfare losses of site closures are larger and more temporally reliable for individuals who take trips to closed sites than the whole sample. Given that the economic value of hunting in the targeted area is consistently high, incentive programs targeted at areas with high CWD prevalence could be effective in engaging the target hunter populations in CWD control. With results from the model and reliability tests, the key finding of largely temporally reliable coefficient and welfare estimates in this study

gives us confidence in using CB data in recreation demand models and policy evaluation and applying them to KT models.

This study provides insights for studies within and beyond recreation demand and policy implications. For researchers who are interested in conducting reliability tests, either with recreation demand models, or with RP, CB, RP-CB data, we provide an extended reliability test framework that considers coefficient estimates, welfare estimates, cross-price elasticities, as well as likelihood ratio tests and non-parametric tests. We show that estimates with CB data are largely temporally reliable and therefore can be used to examine the convergent validity of RP and CB data that are collected in different time periods. Reliable estimates from KT models with CB data suggest that CB data could be combined with RP data to add variation in complicated recreation demand models. Moving beyond the recreation demand literature, our reliability estimates with CB data add confidence to studies in consumer behavior (e.g., Yang et al. 2020), transportation/energy use (e.g., Shin et al. 2012; Ahn, Jeong, and Kim 2008) that collect CB data with discrete-continuous characteristics. Moreover, reliable welfare estimates indicate that researchers and policy makers could rely on CB data to understand costs and benefits of proposed policy programs beforehand. For wildlife managers in Alberta, although with distinct samples of recreational hunters, we show that random samples of hunters' opinions and attitudes towards CWD management programs are likely stable over time without external shocks. Therefore, they could use the findings with information collected from hunters at one time point for future policy design.

This study also has limitations that could be addressed in future work. First, we do not capture the unobserved heterogeneity of the samples in model estimation. The KT model in this study could be extended to incorporate unobserved heterogeneity in a latent class or random parameter KT

models (Lloyd-Smith 2021) that ideally needs more observations. Another source of unobserved heterogeneity comes from hunters' risk perceptions as those who stopped hunting in sampling areas due to perceived high CWD risks were not captured in our samples. The unobserved heterogeneity could be accounted in both sampling/data collection and analysis steps in future work. Second, our findings are based on three years of data rather than a longer time period. As the time intervals could affect the reliability findings (Yi and Herriges 2017; Ji, Keiser, and Kling 2020), one should be cautious when generalizing our findings to a longer time period. Researchers could further examine how time intervals affect temporal reliability of estimates from recreation demand models in a meta-analysis once more studies have examined this question.

Acknowledgement

We would like to thank two anonymous reviewers, the editor, Ian Bateman, Ellen Goddard, Sandeep Mohapatra, Maik Kecinski, and Dana Andersen, and participants at the 2021 EAERE Annual Conference for their valuable comments and suggestions on this paper. We would like to thank John Pattison-Williams, Patrick Lloyd-Smith, Margo Pybus, Anne Hubbs, and Stuart Nadeau for their help with this research. This research is funded by Genome Canada, the Alberta Prion Research Institute, and Alberta Agriculture and Forestry through Genome Alberta in support of the Systems Biology and Molecular Ecology of Chronic Wasting Disease project and the Alberta Prion Research Institute project PEX 18007.

References

Abbott, Joshua K, and Eli P Fenichel. 2013. “Anticipating Adaptation: a Mechanistic Approach for Linking Policy and Stock Status to Recreational Angler Behavior.” *Canadian Journal of Fisheries & Aquatic Sciences* 70 (8): 1190–1208. <https://doi.org/10.1139/cjfas-2012-0517>

Abbott, Joshua K, Patrick Lloyd-Smith, Daniel Willard, and Wiktor Adamowicz. 2018. “Status-quo management of marine recreational fisheries undermines angler welfare. *Proceedings of the National Academy of Sciences* 115 (36): 8948-53. <https://doi.org/10.1073/pnas.1809549115>

Adamowicz, Wiktor, J Louviere, and M Williams. 1994. “Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities.” *Journal of Environmental Economics and Management* 26 (3): 271–92. <https://doi.org/10.1006/jeem.1994.1017>

Ahn, Jiwoon, Gicheol Jeong, and Yeonbae Kim. 2008. “A forecast of household ownership and use of alternative fuel vehicles: A multiple discrete-continuous choice approach.” *Energy Economics* 30 (5): 2091–2104. <https://doi.org/10.1016/j.eneco.2007.10.003>

Bateman, Ian J., Roy Brouwer, S. Ferrini et al. 2011. “Making Benefit Transfers Work: Deriving and Testing Principles for Value Transfers for Similar and Dissimilar Sites Using a Case Study of the Non-Market Benefits of Water Quality Improvements Across Europe.” *Environmental and Resource Economics* 50, 365–387. <https://doi-org.udel.idm.oclc.org/10.1007/s10640-011-9476-8>

Bertram, Christine, Heini Ahtiainen, Jürgen Meyerhoff, Kristine Pakalniete, Eija Pouta, and Katrin Rehdanz. 2020. “Contingent Behavior and Asymmetric Preferences for Baltic Sea Coastal Recreation.” *Environmental and Resource Economics* 75 (1): 49–78. <https://doi.org/10.1007/s10640-019-00388-x>

Bhat, Chandra R. 2008. “The Multiple Discrete-Continuous Extreme Value (MDCEV) Model: Role of Utility Function Parameters, Identification Considerations, and Model Extensions.” *Transportation Research Part B: Methodological* 42 (3): 274–303. <https://doi.org/10.1016/j.trb.2007.06.002>

Bishop, Richard C. and Kevin J. Boyle. 2019. “Reliability and Validity in Nonmarket Valuation.” *Environmental and Resource Economics* 72: 559–582.

Bishop, Richard C., and Kevin J. Boyle. 2017. “Reliability and Validity in Nonmarket Valuation.” In *A Primer on Nonmarket Valuation*, ed. Patricia A Champ, Kevin J Boyle, and Thomas C Brown, 463–97. Dordrecht: Springer Netherlands.

Bockstael, Nancy E., and Kenneth E. McConnell. 2007. *Environmental and Resource Valuation with Revealed Preferences : A Theoretical Guide to Empirical Models*. The economics of non-market goods and resources: v. 7. Dordrecht : Springer, 2007.

Conner, Mary M., Michael W. Miller, Michael R. Ebinger, and Kenneth P. Burnham. 2007. “A Meta-Baci Approach For Evaluating Management Intervention on Chronic Wasting Disease in Mule Deer.” *Ecological Applications* 17 (1): 140–53. [https://doi.org/10.1890/1051-0761\(2007\)017\[0140:AMAFEM\]2.0.CO](https://doi.org/10.1890/1051-0761(2007)017[0140:AMAFEM]2.0.CO)

Cooney, Erin E., and Robert H. Holsman. 2010. “Influences on Hunter Support for Deer Herd Reduction as a Chronic Wasting Disease (CWD) Management Strategy.” *Human Dimensions of Wildlife* 15 (3): 194–207. <https://doi.org/10.1080/10871201003598785>

DeVivo, Melia T., David R. Edmunds, Matthew J. Kauffman, Brant A. Schumaker, Justin Binfet, Terry J. Kreeger, Bryan J. Richards, Hermann M. Schätzl, and Todd E. Cornish. 2017. “Endemic chronic wasting disease causes mule deer population decline in Wyoming.” *PLOS ONE* 12 (10):

e0186512. <https://doi.org/10.1371/journal.pone.0186512>

Englin, Jeffrey, and Trudy Ann Cameron. 1996. “Augmenting travel cost models with contingent behavior data Poisson Regression Analyses with Individual Panel Data.” *Environmental and Resource Economics* 7 (2): 133–47. <https://doi.org/10.1007/BF00699288>

Grijalva, Therese C., Robert P. Berrens, Alok K. Bohara, and W. Douglass Shaw. 2002. “Testing the Validity of Contingent Behavior Trip Responses.” *American Journal of Agricultural Economics VO - 84* (2): 401. <https://www.jstor.org/stable/1244961>

von Haefen, Roger H., and Daniel J. Phaneuf. 2005. “Kuhn-Tucker Demand System Approaches to Non-market Valuation.” In *Applications of Simulation Methods in Environmental and Resource Economics*, ed. Riccardo Scarpa and Anna Alberini, 135–57. U AZ and Stanford U: The Economics of Non-Market Goods and Resources series, vol. 6.

von Haefen, Roger H., and Daniel J. Phaneuf. 2008. “Identifying demand parameters in the presence of unobservables: A combined revealed and stated preference approach.” *Journal of Environmental Economics and Management* 56 (1): 19–32. <https://doi.org/10.1016/j.jeem.2008.01.002>

Hensher, David, Jordan Louviere, and Joffre Swait. 1998. “Combining sources of preference data.” *Journal of Econometrics* 89 (1): 197–221

Holsman, Robert H., and Jordan Petchenik. 2006. “Predicting Deer Hunter Harvest Behavior in Wisconsin’s Chronic Wasting Disease Eradication Zone.” *Human Dimensions of Wildlife* 11 (3): 177–89. <https://doi.org/10.1080/10871200600669916>

Holsman, Robert H., Jordan Petchenik, and Erin E. Cooney. 2010. “CWD After “the Fire”: Six

Reasons Why Hunters Resisted Wisconsin's Eradication Effort." *Human Dimensions of Wildlife* 15 (3): 180–93. <https://doi.org/10.1080/10871201003718029>

Jeon, Yongsik, and Joseph A. Herriges. 2010. "Convergent Validity of Contingent Behavior Responses in Models of Recreation Demand." *Environmental and Resource Economics* 45 (2): 223–50. <https://doi.org/10.1007/s10640-009-9313-5>

Ji, Yongjie, David A. Keiser, and Catherine L. Kling. 2020. "Temporal Reliability of Welfare Estimates from Revealed Preferences." *Journal of the Association of Environmental and Resource Economists* 7 (4): 659–86. <https://doi.org/10.1086/708662>

Lloyd-Smith, Patrick. 2018. "A new approach to calculating welfare measures in Kuhn-Tucker demand models." *Journal of Choice Modelling* 26: 19–27. <https://doi.org/10.1016/j.jocm.2017.12.002>

Lloyd-Smith, Patrick. 2021. "Kuhn-Tucker and Multiple Discrete-Continuous Extreme Value Model Estimation and Simulation in R: The rmdcev Package." *The R Journal* 12(2): 251.

Lloyd-Smith, Patrick. 2022. "The Economic Benefits of Recreation in Canada." *Canadian Journal of Economics*. /Revue canadienne d'économie n/a (n/a).

Lloyd-Smith, Patrick, Joshua K. Abbott, Wiktor Adamowicz, and Daniel Willard. 2020. "Intertemporal Substitution in Travel Cost Models with Seasonal Time Constraints." *Land Economics* 96 (3): 399–417. <https://doi.org/10.3368/le.96.3.399>

Lupi, Frank, Daniel J. Phaneuf, and Roger H. von Haefen. 2020. "Best Practices for Implementing Recreation Demand Models." *Review of Environmental Economics and Policy* 14 (2): 302–23. <https://doi.org/10.1093/reep/reaa007>

Krinsky, Itzhak, and A. Leslie Robb. 1986. "On Approximating the Statistical Properties of Elasticities." *The Review of Economics and Statistics* 68 (4): 715–19.

Murdock, Jennifer. 2006. "Handling unobserved site characteristics in random utility models of recreation demand." *Journal of Environmental Economics and Management* 51 (1): 1–25. <https://doi.org/10.1016/j.jeem.2005.04.003>

Nobel, Anne, Sebastien Lizin, Nele Witters, Francois Rineau, and Robert Malina. 2020. "The impact of wildfires on the recreational value of heathland: A discrete factor approach with adjustment for on-site sampling." *Journal of Environmental Economics and Management* 101: 102317. <https://doi.org/10.1016/j.jeem.2020.102317>

Pattison-Williams, John K., Lusi Xie, W.L. (Vic) Adamowicz, Margo Pybus, and Anne Hubbs. 2020. "An empirical analysis of hunter response to chronic wasting disease in Alberta." *Human Dimensions of Wildlife*: 1–15. <https://doi.org/10.1111/j.1744-7976.1999.tb00388.x>

Phaneuf, Daniel J., and V. Kerry Smith. 2005. "Chapter 15 Recreation Demand Models." In *Handbook of Environmental Economics*, ed. K.-G. M. A. and J. R. Vincent, 2:671–761. Elsevier B.V.

Swait, Joffre, and Jordan Louviere. 1993. "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models." *Journal of Marketing Research* 30 (3): 305–14.

Shin, Jungwoo, Junhee Hong, Gicheol Jeong, and Jongsu Lee. 2012. "Impact of electric vehicles on existing car usage: A mixed multiple discrete–continuous extreme value model approach." *Transportation Research Part D: Transport and Environment* 17 (2): 138–44. <https://doi.org/10.1016/j.trd.2011.10.004>

Vaske, Jerry J., and Katie M. Lyon. 2011. "CWD Prevalence, Perceived Human Health Risks, and State Influences on Deer Hunting Participation." *Risk Analysis* 31 (3): 488–96. <https://doi.org/10.1111/j.1539-6924.2010.01514.x>

Vaske, Jerry J., and Craig A. Miller. 2019. "Deer hunters' disease risk sensitivity over time." *Human Dimensions of Wildlife* 24 (3): 217–30. <https://doi.org/10.1080/10871209.2019.1587650>

Western Association of Fish and Wildlife Agencies. 2017. "Recommendations for Adaptive Management of Chronic Wasting Disease in the West." Edmonton, Alberta Canada and Fort Collins, Colorado, USA.

Whitehead, John C., Daniel J. Phaneuf, Christopher F. Dumas, Jim Herstine, Jeffery Hill, and Bob Buerger. 2010. "Convergent Validity of Revealed and Stated Recreation Behavior with Quality Change: A Comparison of Multiple and Single Site Demands." *Environmental and Resource Economics* 45 (1): 91–112. <https://doi.org/10.1007/s10640-009-9307-3>

Xie, Lusi, Wiktor L. Adamowicz, and Patrick Lloyd-Smith. 2020. "Spatial and Temporal Responses to Incentives: An Application to Wildlife Disease Management." Working paper. University of Alberta, Edmonton.

Yang, Ou, Peter Sivey, Andrea M de Silva, and Anthony Scott. 2020. "Parents' Demand for Sugar Sweetened Beverages for Their Pre-School Children: Evidence from a Stated-Preference Experiment." *American Journal of Agricultural Economics* 102 (2): 480–504. <https://doi.org/10.1002/ajae.12033>

Yi, Dong Gyu, and Joseph A. Herriges. 2017. "Convergent validity and the time consistency of preferences: Evidence from the Iowa Lakes recreation demand project." *Land Economics* 93 (2): 269–91. <https://doi.org/10.3368/le.93.2.269>

Zimmer, Natalie M.P., Peter C. Boxall, and Wiktor L. Adamowicz. 2012. “The Impacts of Chronic Wasting Disease and Its Management on Recreational Hunters.” *Canadian Journal of Agricultural Economics* 60 (1): 71–92. <https://doi.org/10.1111/j.1744-7976.2011.01232.x>

Tables

Table 1. Contingent Behavior Scenarios

Scenario	Description	Eligible areas^a	Season Length	Material Incentives
October season expansion	Extend the hunting season from the entire month of November to include the last week of October	Sampling areas (65 sites)	37 days	NA
December season expansion	Extend the hunting season from the entire month of November to include the first 17 days of December	Sampling areas (65 sites)	47 days	1 extra tag if hunting in December ^b
Extra tags	Add one extra hunting tag in November	High CWD prevalence areas (11 sites)	30 days	1 extra tag
Gift cards	Offer gift cards from a popular hunting store for animals harvested in November	High CWD prevalence areas (11 sites)	30 days	1 gift card (valued at \$30 or \$50)

Note:

^a Appendix Figure A2 provides maps of eligible areas under each scenario.

^b The extra tag in December season expansion scenario only applies to the extended season. This is to make the scenario more feasible: as the number of animals harvested is restricted by hunting tags in Alberta, recreational hunters would not have taken more trips in December if they already used up the hunting tags in November.

Table 2 Mean Values of Site Attributes, Contingent Behavior Scenario and Socio-demographic Variables

Variable^a	Description	2018	2019	2020
<i>Site attributes</i>				
CWD	Chronic Wasting Disease (CWD) prevalence rate (%) available from hunting season of 2016, 2017, and 2018 ^b	1.927	3.680	5.990
<i>CB scenario dummy variables^c</i>				
October scenario	Dummy variable if the October season expansion scenario is proposed in eligible areas	0.263	0.289	0.286
December scenario	Dummy variable if the December season expansion scenario is proposed in eligible areas	0.292	0.318	0.304
Extra tags scenario	Dummy variable if the extra tags scenario is proposed in eligible areas	0.023	0.023	0.019
Gift cards scenario	Dummy variable if the gift cards scenario is proposed in eligible areas	0.022	0.022	0.021
Extended season	Dummy variable if the trip is taken during the extended hunting seasons	0.336	0.352	0.358
<i>Socio-demographic variables</i>				
College	Dummy variable if hold a college degree	0.328	0.430	0.337
Urban	Dummy variable if live in urban area (20,000 people or more)	0.501	0.531	0.475
Children	Dummy variable if children under 12 in household	0.240	0.238	0.216
Years of hunt	Years of hunting experience	25.033	28.769	28.491
Income	Annual household income	99,202	105,232	104,012
<i>Travel cost</i>	Travel cost in 2017 Canadian dollars	270.365	316.546	284.529
N	Number of respondents	636	330	873

Note:

^a Not all variables are balanced across years according to two-sample t-tests and joint orthogonality tests. However, most socio-demographic variables such as age and income have similar distributions across years.

^b The surveys were conducted in 2018, 2019, and 2020 to collect RP and CB trip data in the previous hunting season in 2017, 2018, and 2019. However, as CWD testing results came after the hunting season, hunters only had CWD information from the previous season (i.e., 2016, 2017, and 2018) when making trip decisions in 2017, 2018, and 2019.

^c Dummy variables of October scenario, December scenario, Extra tags scenario, Gift cards scenario are 0 when these scenarios do not apply (e.g., either in ineligible areas, or an individual did not receive the scenarios).

Table 3. Selected Kuhn-Tucker Model Parameter Estimates

	2018	2019	2020
Baseline marginal utility parameters (β_j)			
CWD	-7.188 (8.104)	-10.291 (8.966)	-11.191 (7.106)
October scenario	0.204*** (0.062)	0.075 (0.074)	0.087* (0.048)
December scenario	0.308*** (0.062)	0.162*** (0.061)	0.154*** (0.047)
Extra tags scenario	0.614*** (0.075)	0.523*** (0.082)	0.583*** (0.061)
Gift cards scenario	0.317*** (0.088)	0.182** (0.089)	0.366*** (0.056)
Extended season	-0.087** (0.044)	-0.160*** (0.052)	-0.106*** (0.031)
College	-0.024 (0.054)	-0.018 (0.051)	-0.028 (0.033)
Urban	-0.134 (0.061)	-0.112 (0.065)	-0.067* (0.037)
Children	0.067** (0.044)	0.073* (0.063)	0.049* (0.026)
Years of hunt	-0.009 (0.018)	0.020 (0.019)	-0.015 (0.011)
Satiation parameter (α)	0.204*** (0.031)	0.202*** (0.035)	0.265*** (0.020)
Scale parameter (σ)	0.560*** (0.011)	0.545*** (0.013)	0.535*** (0.008)
Number of observations	1285	883	2234
Number of respondents	636	330	873
Log-likelihood	-12843.94	-8503.28	-21005.30

Note:

This table reports selected estimates for KT model parameters. Standard errors computed using 50 multivariate normal draws are in parenthesis. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

The CWD variable is in absolute values rather than in percentage. Years of hunt index is scaled as the year of hunting experience divided by 10.

One alternative specific constant (ASC) in β_j and γ_j parameters are estimated for each hunting site. They are not reported here but included in Appendix B.

Table 4. Welfare Estimates (CAD\$) Per Trip of Closing Selected Sites

Per person

(i.e., averaging over all respondents in each survey)

Site	2018			2019			2020		
	Mean	Low	High	Mean	Low	High	Mean	Low	High
148	-2.46	-3.23	-1.89	-3.79	-4.56	-3.00	-2.37	-2.97	-1.97
150	-3.14	-4.49	-2.46	-3.88	-5.23	-2.81	-2.42	-3.18	-2.03
151	-5.17	-6.40	-3.87	-3.97	-5.27	-3.00	-4.24	-4.89	-3.57
152	-2.86	-3.48	-2.25	-3.60	-4.59	-2.70	-2.10	-2.45	-1.82
163	-3.64	-4.55	-2.64	-4.02	-5.46	-2.84	-3.02	-3.59	-2.44
200	-3.38	-4.26	-2.43	-7.40	-9.83	-5.82	-6.02	-6.63	-5.19
202	-3.05	-3.84	-2.38	-3.37	-4.16	-2.54	-3.33	-3.96	-2.79
234	-3.71	-4.41	-2.99	-3.94	-4.87	-2.98	-5.28	-6.06	-4.50
236	-1.63	-2.06	-1.22	-1.50	-2.10	-0.95	-2.10	-2.45	-1.78
728	-2.05	-2.91	-1.43	-0.28	-0.57	-0.05	-0.78	-1.04	-0.51
730	-1.61	-2.19	-0.97	-0.55	-1.09	-0.17	-0.11	-0.23	-0.04

Per participant

(i.e., averaging over respondents who would have taken at least 1 trip to the site in each survey)

Site	2018			2019			2020		
	Mean	Low	High	Mean	Low	High	Mean	Low	High
148	-61	-80	-47	-62	-75	-49	-55	-69	-46
150	-101	-144	-79	-84	-113	-60	-72	-95	-60
151	-98	-121	-73	-100	-133	-76	-91	-105	-77
152	-58	-71	-46	-58	-74	-43	-52	-61	-45
163	-109	-136	-79	-99	-134	-70	-89	-106	-72
200	-89	-112	-64	-96	-128	-76	-91	-100	-78
202	-71	-90	-56	-83	-102	-62	-65	-78	-55
234	-67	-80	-54	-68	-84	-52	-72	-82	-61
236	-42	-53	-31	-74	-103	-46	-54	-63	-46
728	-83	-117	-58	-62	-125	-10	-92	-122	-60
730	-130	-176	-78	-121	-241	-37	-84	-173	-28

Note:

This table reports the average welfare estimates per trip of closing sites (one at a time) with high CWD prevalence and all CB scenarios applied. Appendix C reports the average welfare estimates per trip of closing every site one at a time in choice sets in three years.

Welfare estimates per person are calculated by averaging welfare estimates over the whole sample for each survey. Welfare estimates per participant are calculated by averaging welfare estimates over respondents who would have taken at least 1 trip to corresponding sites in each survey.

Low and high are 95% confidence intervals of the mean estimates calculated from 30 simulations with 50 individual conditional error draws.

Figure Captions

Figure 1. KT Model Estimates: ϕ Parameters of CB Scenarios

Note: Dots represent the mean estimates of parameters. The dashed horizontal line is the zero reference line. Error bars are in capped vertical lines, representing 95% confidence intervals calculated using 50 multivariate normal draws.

Figure 2. Welfare Estimates of Site Closures

Note: Dots represent the average welfare estimates of closing three hunting sites (i.e., Wildlife Management Units, WMUs): WMUs 150 (with high CWD prevalence and all CB scenarios applied), 230 (with CWD presence and only season expansion scenarios applied) and 501 (without CWD presence and no CB scenarios applied) respectively. The dashed horizontal line is the zero reference line. Error bars are in capped vertical lines, representing 95% confidence intervals calculated from 30 simulations with 50 individual conditional error draws.

Figure 3. Percentage of Temporally Reliable Coefficient and Welfare Estimates

Endnotes

¹In Bhat (2008) and other literature in transportation, the Kuhn-Tucker (KT) model with a multiple discrete-continuous extreme value (MDCEV) specification is usually called “MDCEV”. In this paper, we call it as Kuhn-Tucker model to follow the literature in environmental economics. However, one should note our KT model specification is different from another KT model specification by von Haefen and Phaneuf (2005).

²As pointed out by a reviewer, the advancement of different software packages has addressed the computational challenges associated with the KT models in the early days. However, KT models have not gained much popularity among environmental economists, and it is an open question why they are less popular compared to discrete choice models.

³Ji et al. (2020) is the closest study to ours as they examine the temporal reliability of estimates from KT models in the von Haefen and Phaneuf (2005) Linear Expenditure System specification with RP data as a robustness check in their analysis of DCM reliability.

⁴<https://www.usgs.gov/centers/nwhc/science/expanding-distribution-chronic-wasting-disease>

⁵<https://www.inspection.gc.ca/animal-health/terrestrial-animals/diseases/reportable/cwd/fact-sheet/eng/1330189947852/1330190096558>

⁶<https://open.alberta.ca/dataset/d850792e-cd0c-4bb5-b10e-8e84eae0d764/resource/3b69d983-8dee-45bd-8924-c52d8d2707db/download/cwd-positivedeer-infosheet-sep2018.pdf>

⁷For example, if an individual received the extra tag scenario, the scenario dummy variable is 1 only for the 11 sites. The dummy variable is 0 in all ineligible sites regardless whether an individual received the scenario or not.

⁸The functional form is chosen over other profiles in Bhat (2008) based on model fit statistics of log-likelihood values.

⁹Differences in the choice set and CB scenarios with Xie, Adamowicz, and Lloyd-Smith (2020) are also explain the reason why we define CB dummy variables of October or December scenarios differently: to capture spatial variation of CB scenarios.

¹⁰The standard errors can also be computed with the delta method. However, we use multivariate normal draws for consistency as they are required to use for welfare simulation (Lloyd-Smith 2021).

¹¹We thank an anonymous reviewer for this suggestion.

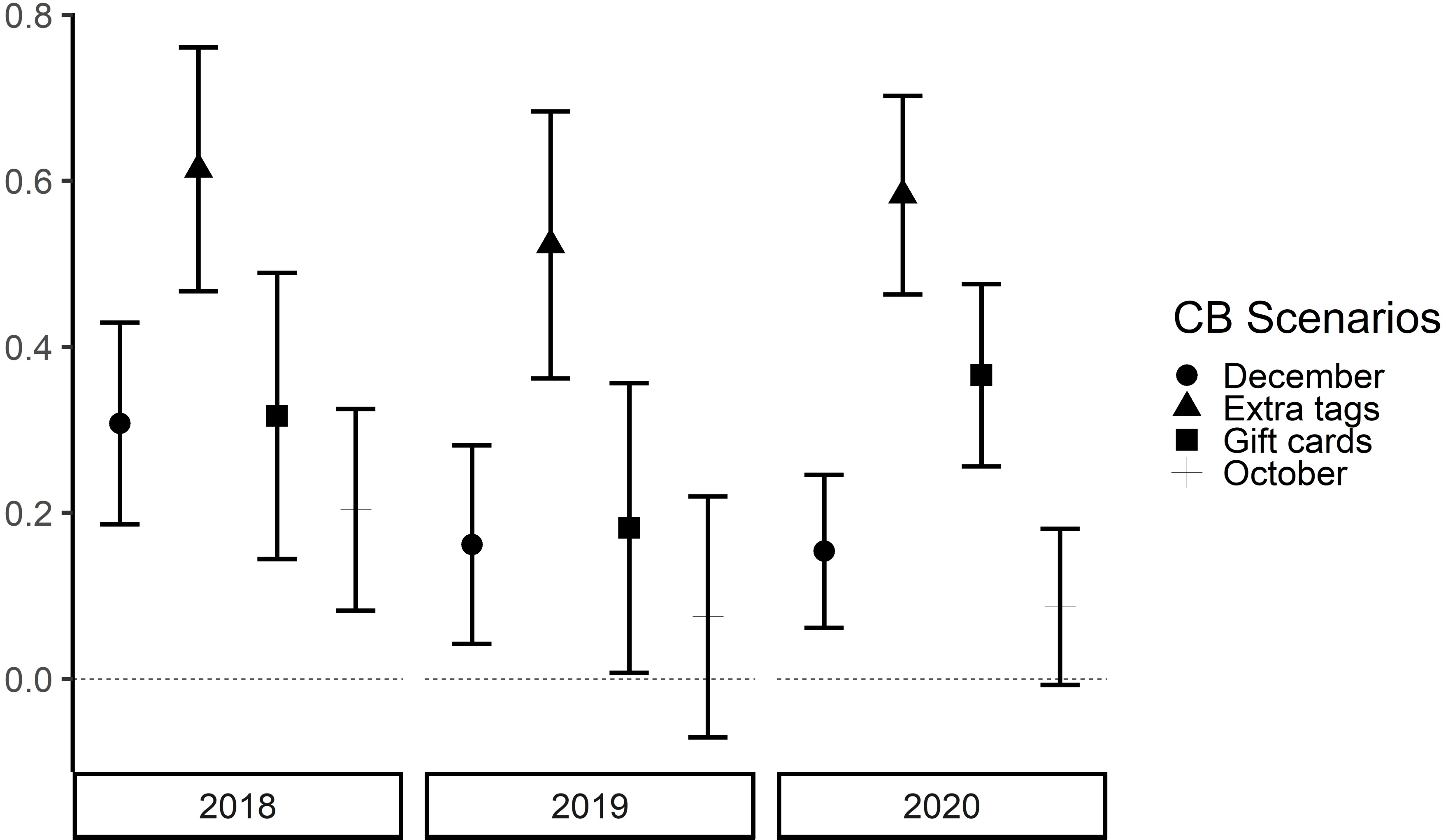
¹²See <https://github.com/LusiXie/CB-reliability-LE> for replication R code. We thank an anonymous reviewer for this suggestion.

¹³Different from a repeated choice model where one ASC for a site is omitted as the reference category, we include one ASC for each site and use the numeraire good as the reference category. Although Lloyd-Smith (2021) proposes an option to leave out one ASC due to identification concerns, we do not use this option in the final model as the log-likelihood at convergence with all ASCs is larger than the one leaving out one ASC.

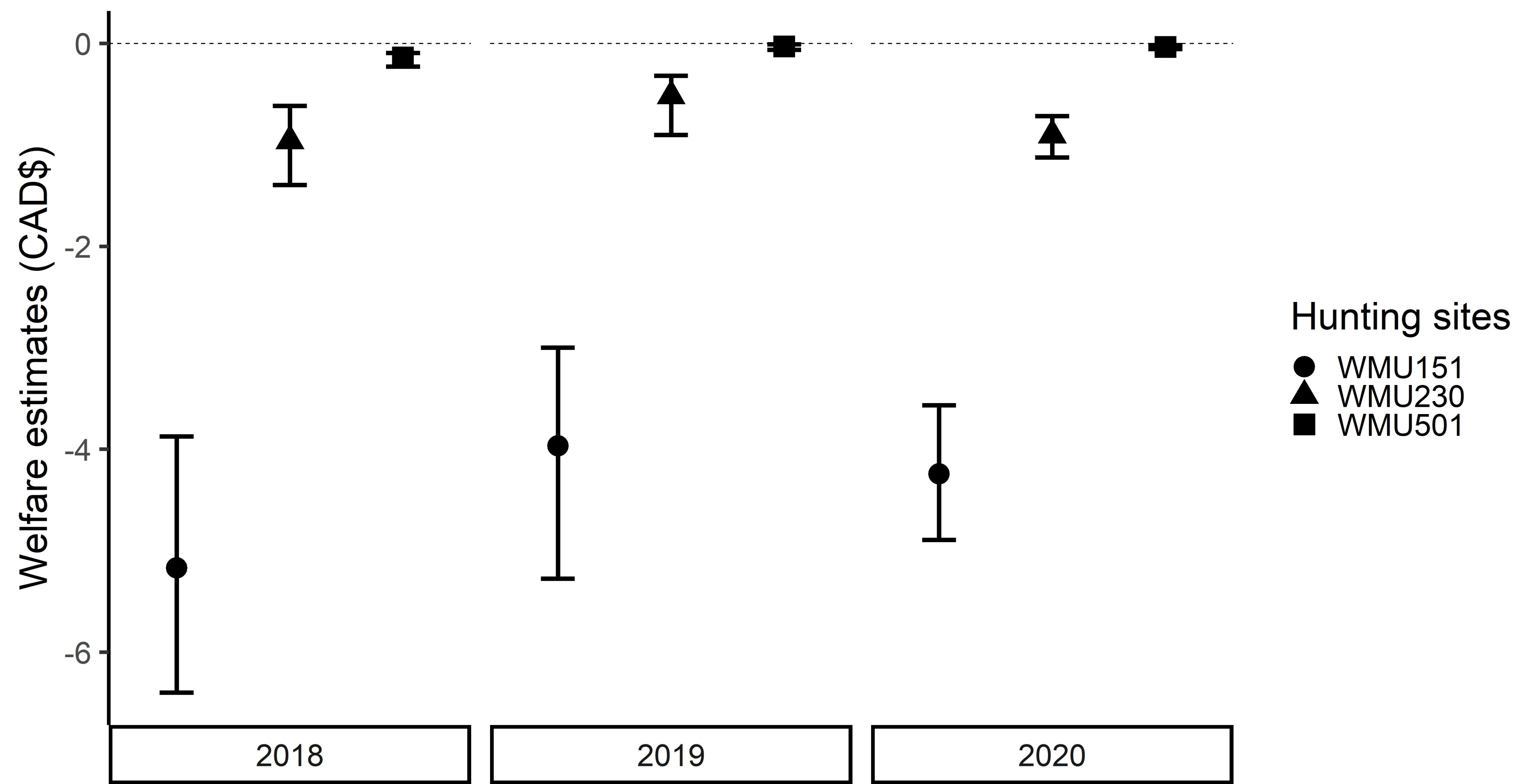
¹⁴As suggested by a reviewer, we also estimated the models with categorical CWD measures as a robustness check. The results were very similar to the results from continuous CWD measures regarding the magnitudes, signs, and significance levels of the coefficient estimates.

¹⁵We thank an anonymous reviewer for this suggestion.

Psi parameter estimates



Per Person



Per Participant

